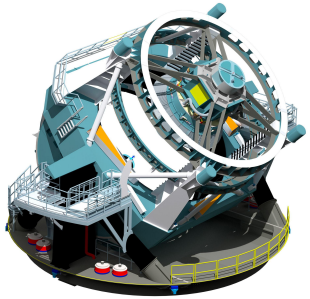
LSST-France, LPNHE, 23 November 2021

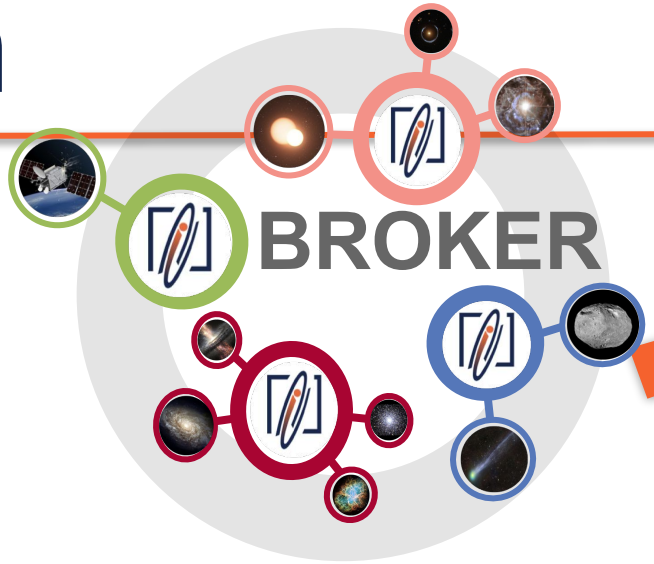
Emille Ishida, Julien Peloton and Anais Moller
on behalf of the Fink Team

Data path

XX deg2 every ~30 seconds
down to mag ~24



10 million alerts
per night…

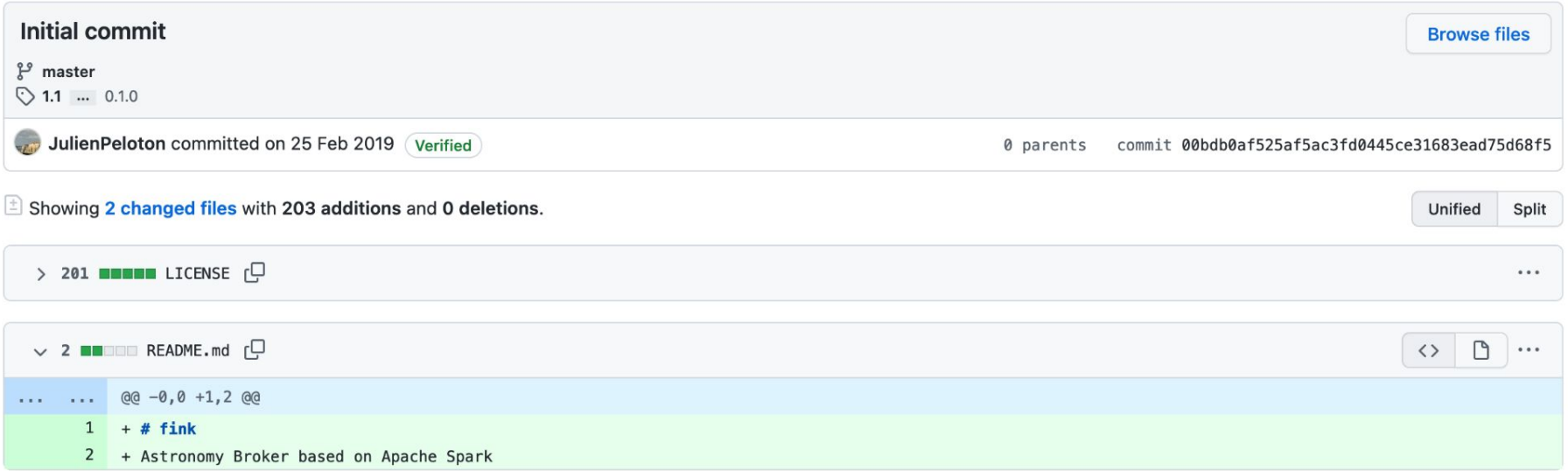


Machine learning
Catalog association
Streams join

We would like the interesting ones ..

Back in 2019...

Following an invited seminar at LPC-Clermont ...



… and we are selected!

- Broker selection is done -- Fink selected!
  - *7 laureates + two 2nd layer brokers*
  - *Deployment at CC-IN2P3 for 2022*

Thanks all, and congratulations for all the work and support over the last 2 years

Fink in the world

Fink in Europe

Since **Sep/2021**, we started coordinating efforts among the **European brokers** in order to:

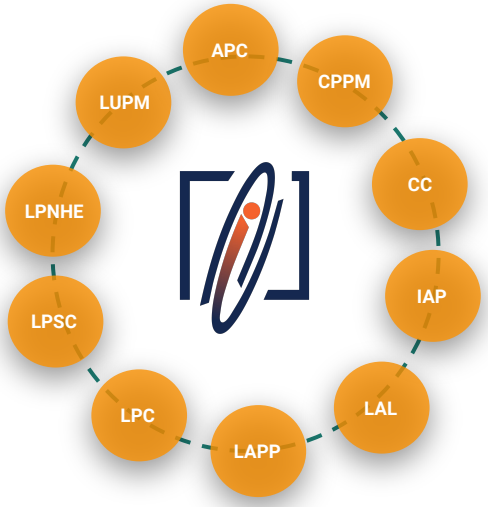
- Share expertise
- Avoid double effort (stamps)
- Agree on conventions, formats, namings, etc.
- Start direct dialog with CDS

<u>Current status:</u>

- Bi-monthly meetings
- CDS staff involved from early stages
- **Lasair will incorporate Fink Early SN scores**
- Fink plans to I integrate Sherlock cross-matches

Fink community

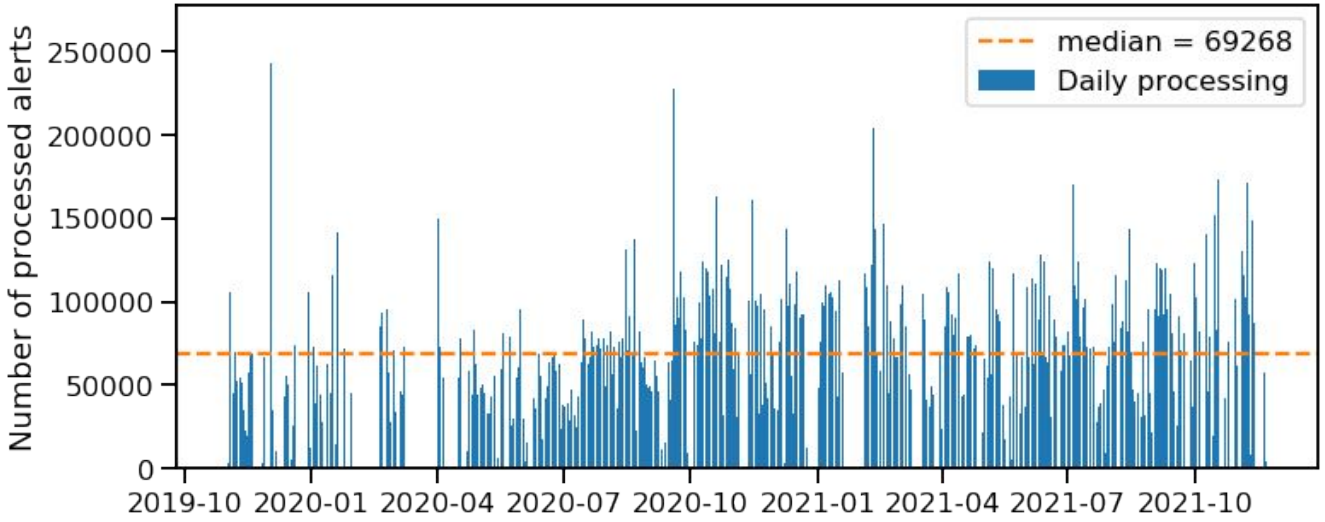
- 52 registered (~20% from abroad: Australia, Czech Republic, Russia, Switzerland, UK, USA)
- In France: > 10 labs (IN2P3 and INSU)
- Monthly general meetings
- Organization: management plan, code of conduct
  - Soon to come: scientific advisory board



ZTF public data stream

Processing public ZTF alerts since 11/2019

- 120,000,000 alerts received
- 40,000,000 alerts processed



Fink in numbers

Data streams (571 nights)

- 120,000,000 alerts received since 2019
- 40,000,000 alerts processed since 2019

Cloud computing facilities

- 24/7 processing at VirtualData*
- Dedicated Spark, Kafka, HBase clusters, with Kubernetes deployment
- ~10 TB of raw+enriched data (HDFS)
- Hundreds of request per day on the web servers -- 10 ms response for individual object query
- Stream simulator in place (up to 200MB/s, 7 times LSST rate!)

Redundancies and recovering mechanisms at all stages

*When no power outage :D

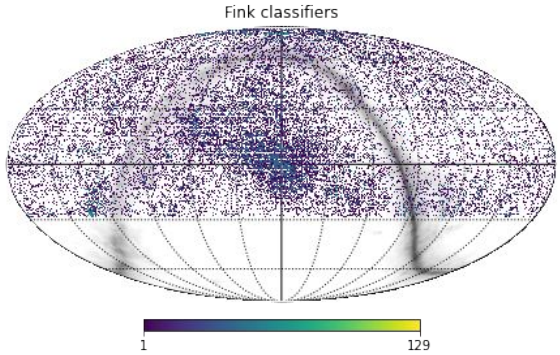
Probing different scales

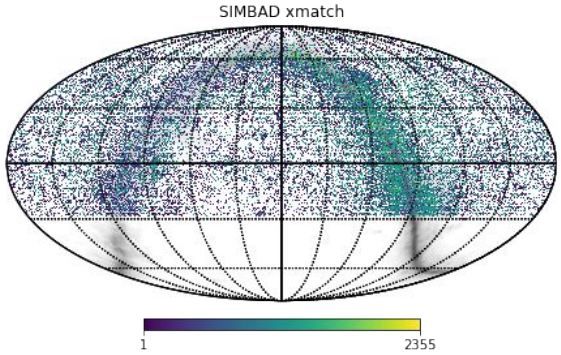
Two main crossmatch:

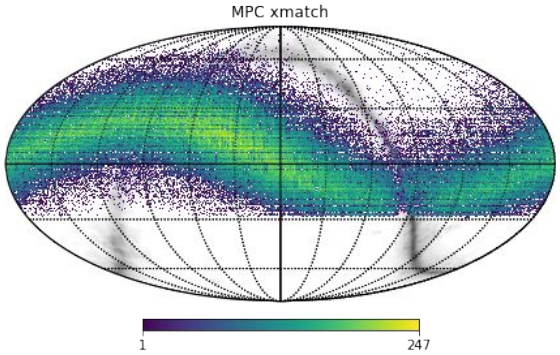
- SIMBAD database, mostly galactic objects
- MPC catalog, mostly objects on the ecliptic

Fink classifiers mostly targeting extragalactic objects

Future: extend our crossmatch to cover extragalactic objects, and classifiers to cover galactic objects



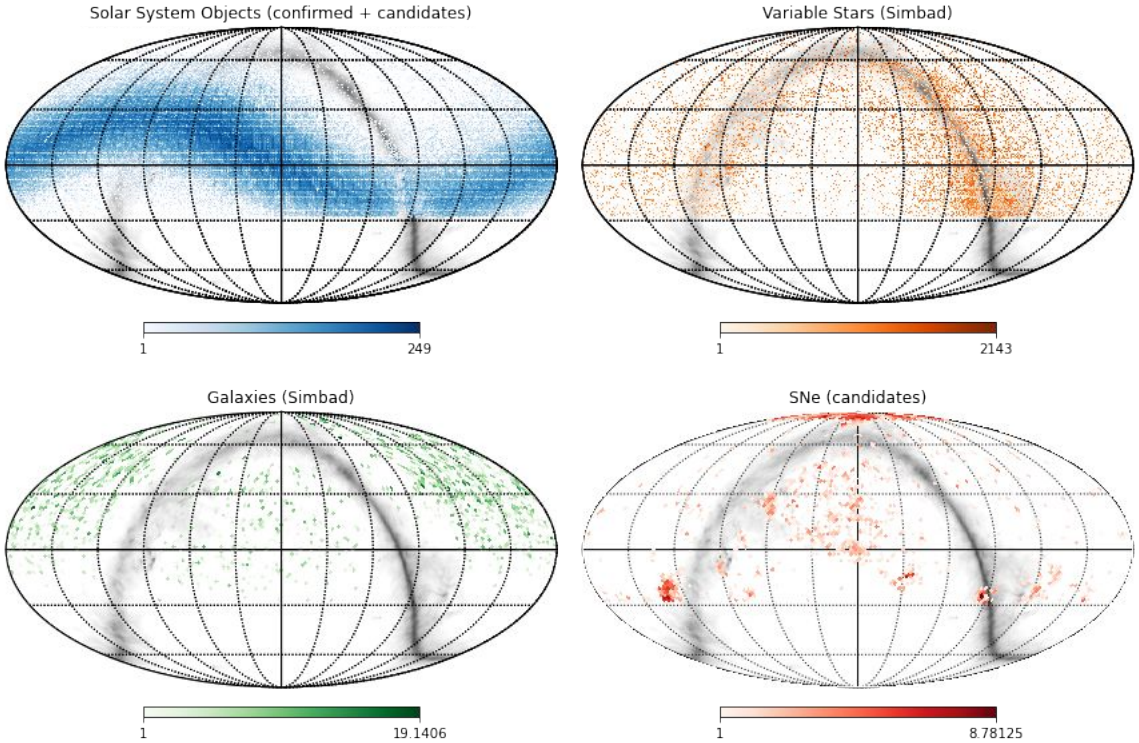




Fink substreams

Every night

- 10,000+ known SSO
- 10,000+ variable Stars
- 100+ SNe candidates
- 100+ SSO candidates
- 10+ early SN Ia candidates
- 10+ satellite debris
- 1+ fast transient candidate (GRB, Kilonovae)
- ...

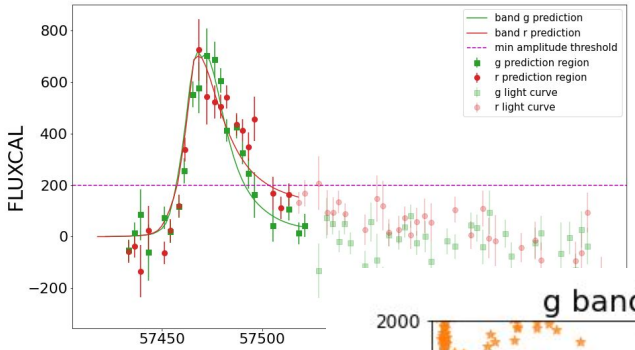


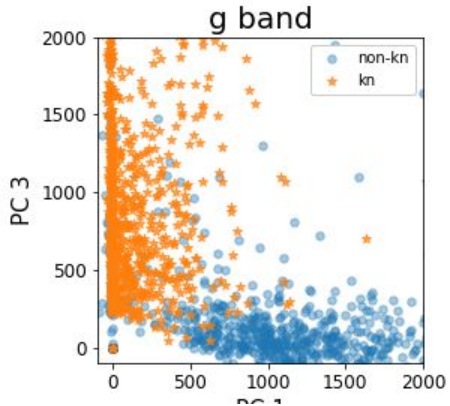
Current project highlights

- **Long** transients (Blodin, Pruzhinskaya)
- AL for **early SN Ia** discovery (Leoni)
- **Anomaly** detection (Pruzhinskaya, Kornilov, Russeil, Beschastnov)
- Detection and classification of **satellite glints & debris** (Karpov)
- Search for **GRB** afterglows (+orphans) (Bregeon, Turpin, Ducoin, Le Montagner)
- Search for **Kilonova** (Biswas, GRANDMA collaboration)
- New techniques to discover **SSO** objects (Le Montagner)
- New database techniques using **graphs** (Hrivnac)

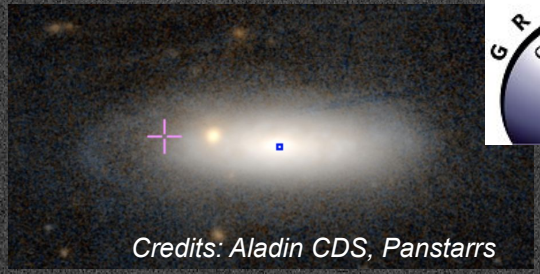
Different colors = different groups

Fink science: KN search





ML based scores, (*Biswas et al., in prep)* + cross-match based filter



Candidates sent to amateur astronomers for additional observations

Credits: D. Boutigny

Fink science: GRB

See Damien and Johan's talks

Search for orphan GRB afterglow (Bregeon+)

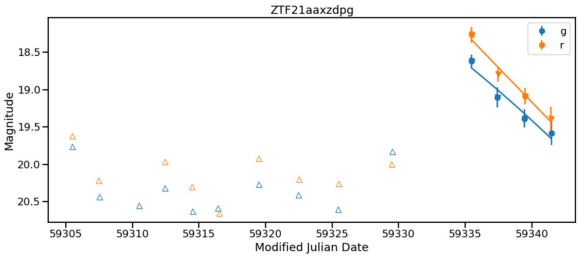
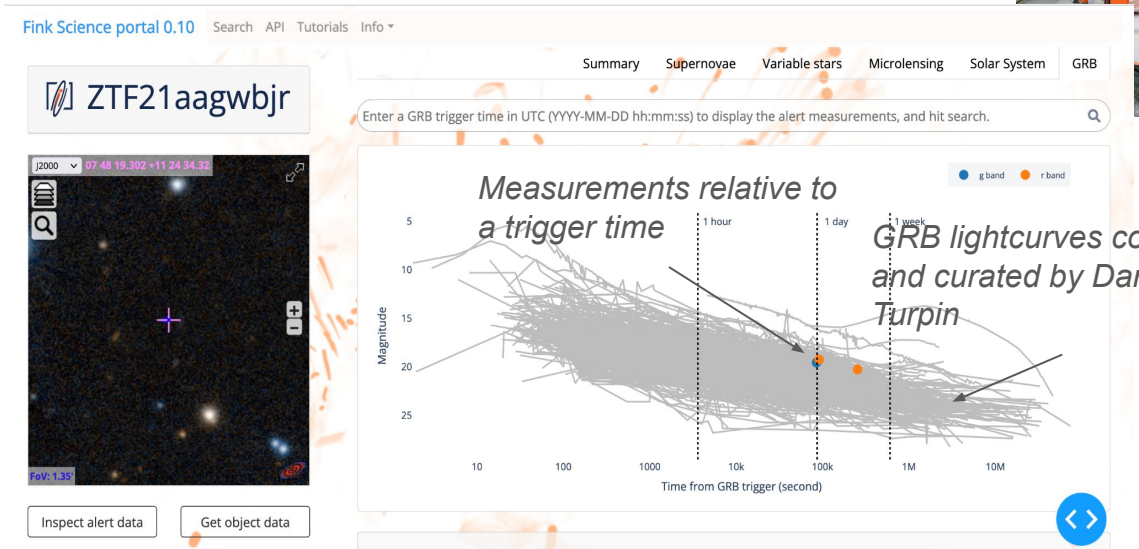
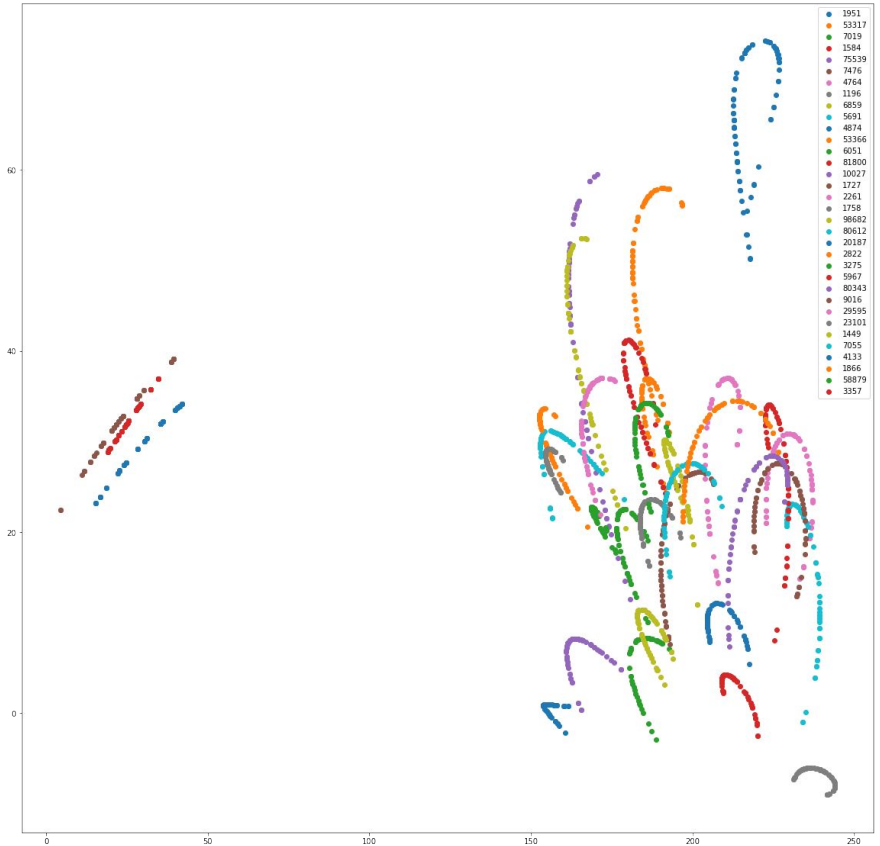


FIGURE 2 – Courbe de lumière de l'objet ZTF21aaxzdpg



Fink science: SSO

Using MPC-labeled objects (~10,000 per night) to train model, reconstruct trajectories from SSO candidates (~100 per night).



Fink Science: EarlySN

Leoni *et al.* 2021 - arXiv:astro-ph:/XXXXX

Fink: early supernovae Ia classification using active learning

M. Leoni[1]*, E. E. O. Ishida[2]**, J. Peloton[1] and A. Möller[2,3]

[1] Université Paris-Saclay, CNRS/IN2P3, IJCLab, 91405 Orsay, France
[2] LPC, Université Clermont Auvergne, CNRS/IN2P3, F-63000 Clermont-Ferrand, France
[3] Centre for Astrophysics and Supercomputing, Swinburne University of Technology, Mail Number H29, PO Box 218, 31122 Hawthorn, VIC, Australia

Received September 15, 1996; accepted March 16, 1997

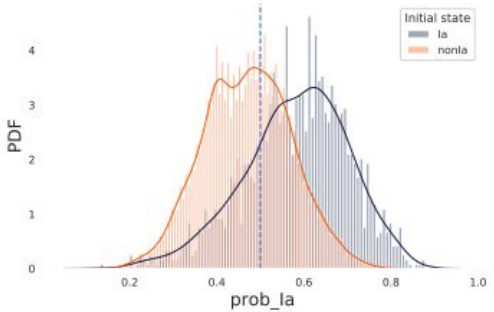
ABSTRACT

Context. The Vera C. Rubin Observatory Legacy Survey of Space and Time (LSST) will produce a continuous stream of alerts made of varying sources in the sky. This data flow will be publicly advertised and distributed to scientists via broker systems such as Fink, whose task is to extract scientific information from the stream. Given the complexity and volume of the data to be generated, LSST is a prime target for machine learning (ML) techniques. One of the most challenging stages of this task is the construction of appropriate training samples which enable learning based on a limited number spectroscopically confirmed objects.
Aims. We describe how the Fink broker early supernova Ia classifier optimizes its ML classifications by employing an active learning (AL) strategy. We demonstrate the feasibility of implementation of such strategies in the current Zwicky Transient Facility (ZTF) public alert data stream.
Methods. We compare the performance of two AL strategies: uncertainty sampling and random sampling. Our pipeline consists of 3 stages: feature extraction, classification and learning strategy. Starting from an initial sample of 10 alerts (5 SN Ia and 5 non-Ia), we let the algorithm identify which alert should be added to the training sample. The system is allowed to evolve through 300 iterations.
Results. Our data set consists of 23 840 alerts from the ZTF with confirmed classification via cross-match with SIMBAD database and the Transient name server (TNS), 1 600 of which were SNe Ia (1 021 unique objects). The data configuration, after the learning cycle was completed, consists of 310 alerts for training and 23 530 for testing. Averaging over 100 realizations, the classifier achieved ~89% purity and ~54% efficiency. From 01/November/2020 to 31/October/2021 Fink has applied its early supernova Ia module to the ZTF stream and communicated promising SN Ia candidates to the TNS. From the 535 spectroscopically classified Fink candidates, 459 (86%) were proven to be SNe Ia.
Conclusions. Our results confirm the effectiveness of active learning strategies for guiding the construction of optimal training samples for astronomical classifiers. It demonstrates in real data that the performance of learning algorithms can be highly improved without the need of extra computational resources or overwhelmingly large training samples. This is, to our knowledge, the first application of AL to real alerts data.

Fink Science: EarlySN

Final model:

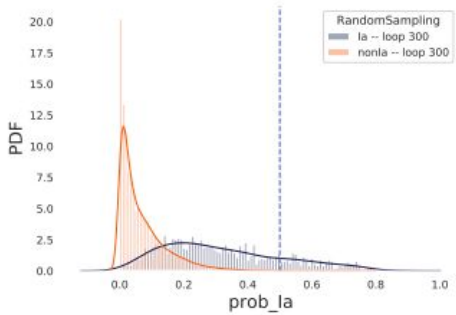
- Random Forest classifier
- **310** alerts used for training (TNS+Simbad)
- **23 530** alerts for testing (TNS+Simbad)
- Efficiency: 0.53, Purity: 0.92

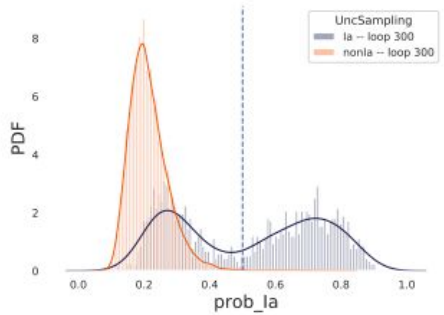


Use only 30 days photometric history in alerts

From November/2020 to October/2021:

- 809 candidates sent to TNS
- 535 spectroscopically confirmed
- 459 classified as Ia (86%)





Leoni et al, 2021 - submitted to A&A

Other ongoing efforts

- CNRS-MITI project application for anomaly detection (Blodin, Pruzinskaya)
- Galaxy zoo connection
- GOTO - proposed module to use the stamps
- Fink-TOM module in the context of LSST-DESC (w/ Cohen-Tanugi)
- Oz-Fink: connection with follow-up facilities in Australia
- ...

Engaging LSST France community

- Science cases which will help us prepare for commissioning
- Possibilities:
  - Build a science module
  - Use the output of the science modules and filters already in place

Fink was designed to allow easy implementation of very specific science needs…

… do not be afraid to be creative!

- Form proposal: *https://fink-broker.org/joining.html*
- Email: *contact@fink-broker.org*
- Twitter: @FinkBroker

Coordination & interoperability

Identifying interesting LSST alerts is only part of the story: we need coordination with other facilities, follow-up resources and existing networks.

- Discussions and work with teams from: SVOM, GRANDMA, CTA, Integral, KM3NET, …
  - Work on GRB (Fermi/SWIFT-ZTF) to prepare SVOM-LSST
    - + recent partnership with ground-based follow-up telescope
  - GRANDMA (network & follow-up for KNe)

We will regularly publicize a prioritized list of targets for each science case that should be followed in order to improve future estimates.

- How to integrate this in the current landscape given the scale?
- How to coordinate with existing follow-up resources (+ToO) and surveys?

Food for thought

- Fink is adding value to ZTF data
  - e. g., EarlySN scores can be used for a variety of science cases
  - Similar classifiers can easily be trained for other transient classes

- Our main goal is to provide support for the LSST-France and European communities

- What are the capabilities you would like to see ?

- Do you foresee any connections with experiments or communities not currently being tackled?

We welcome your feedback at any time: *contact@fink-broker.org*

Extra slides

Fink was selected!

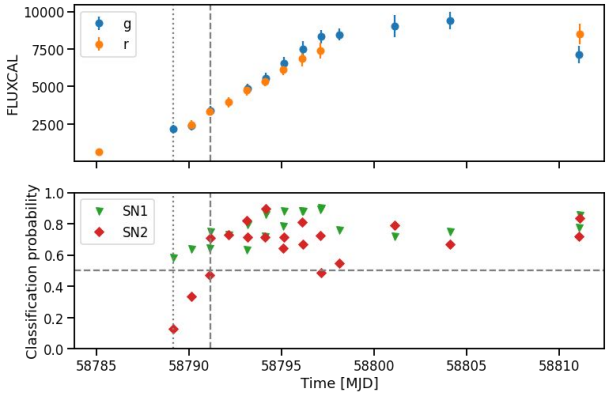
2.5 years ago, we were writing the Letter of Intent for LSST… A lot happened since then!

- Aug 2019: pre-selection by the Rubin Observatory
- Feb 2020: MoU with ZTF-I (renewed for ZTF-II)
- Oct 2020: Organisation of LSST broker workshop part 1
- Nov 2020: Publication of Fink white paper (MNRAS)
- Nov 2020: Fink becomes IN2P3 project (LSST Master project)
- Dec 2020: Submission of Full Proposal to the Rubin Observatory
- Feb 2021: Joint proposal on CFHT with GRANDMA for Kilonovae search
- Apr 2021: Organisation of LSST broker workshop part 2
- May 2021: Citizen science Kilonova-Catcher project with GRANDMA
- Jul 2021: Fink is officially selected to receive the full LSST alert stream

SN classification

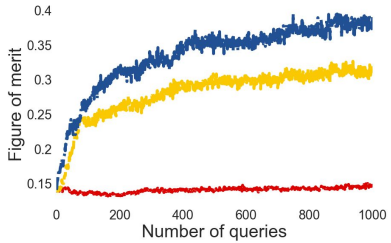
High-accuracies using *Deep Learning*

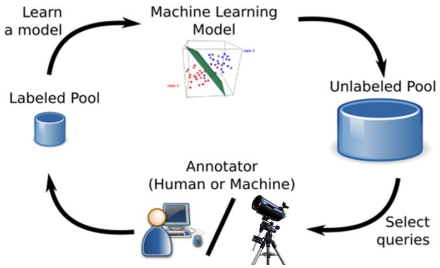
Moller and Boissiere, 2020, MNRAS



- All science modules are built by the community
- Outputs from all science modules will be publicly available
- Hosted at CC-IN2P3: cross-match with LSST image data available for users with data rights
- Distributed Machine Learning

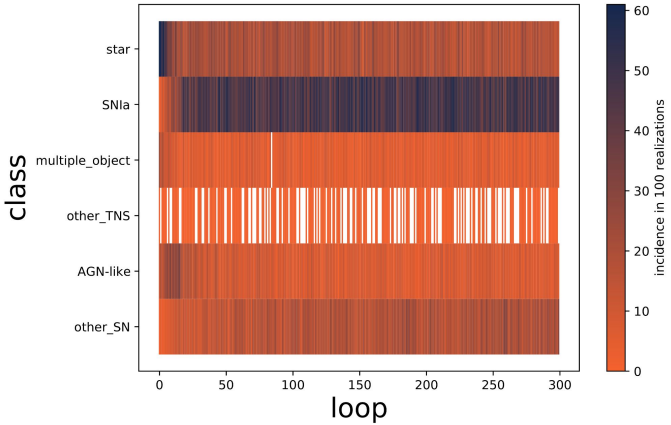
Increasingly more accurate classifications and anomaly scores using *Adaptive Learning*

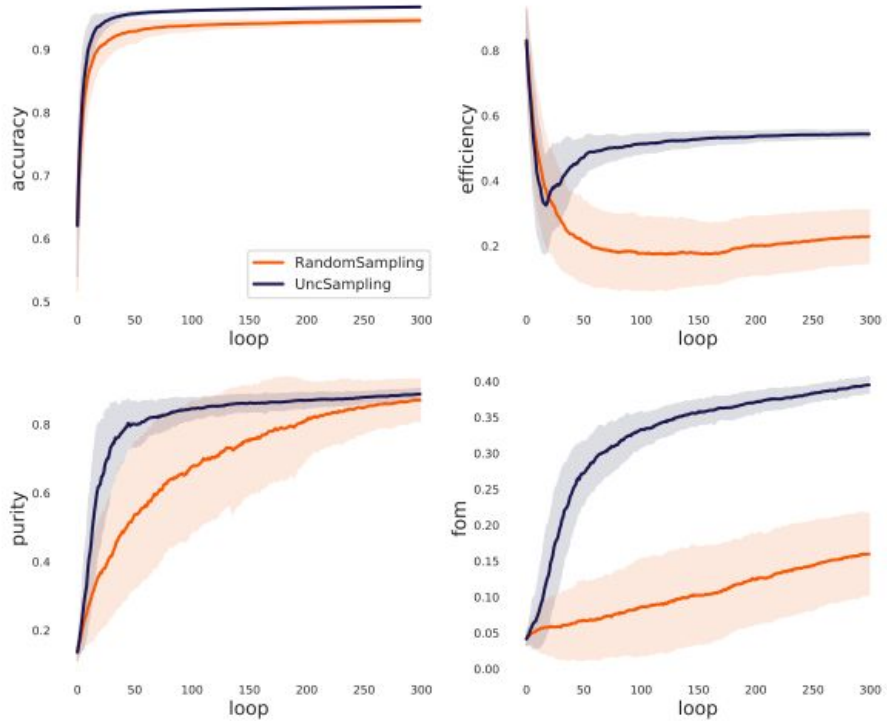




Ishida et al., 2019, MNRAS

Fink Science: EarlySN





Leoni et al, 2021 - submitted to A&A

Fink team

PI: E. Ishida (LPC), A. Möller (Swinburne Uni.), J. Peloton (IJCLab)

30+ members all over EU

Fink members & collaborators

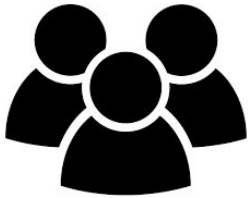
- **CEA/AIM, Université Paris-Saclay**: D. Turpin
- **Centre for Data Intensive Science/MSSL, University College London**: T. Allam Jr.
- **CEICO, Institute of Physics, Czech Academy of Sciences**: S. Karpov
- **CNRS/APC, Université de Paris**: A. Boucaud, A. Coleiro
- **CNRS/CC-IN2P3**: F. Hernandez
- **CNRS/CDS, Observatoire Astronomique de Strasbourg, Université de Strasbourg**: A. Nebot Gomez-Moran
- **CNRS/CPPM, Aix Marseille Université**: D. Dornic, D. Fouchez
- **CNRS/IJCLab, Université Paris-Saclay**: R. Ansari, T. Blaineau, J.-E. Campagne, R. Le Montagner, N. Leroy, M. Moniez, J. Neveu
- **CNRS/IRAP, Université Paul Sabatier**: O. Godet, N. Webb
- **CNRS/LAPP, Université Grenoble-Alpes, Université Savoie Mont Blanc**: D. Boutigny
- **CNRS/LPC, Université Clermont Auvergne**: E. Gangler, P. Gris, E. Russeil
- **CNRS/LPSC, Université Grenoble-Alpes**: J. Bregeon
- **ISDC, Department of Astronomy, University of Geneva**: V. Savchenko
- **Las Cumbres Observatory**: E. Bachelet
- **Lomonosov Moscow State University, Sternberg Astronomical Institute**: M. V. Kornilov, M. V. Pruzhinskaya

Development Team

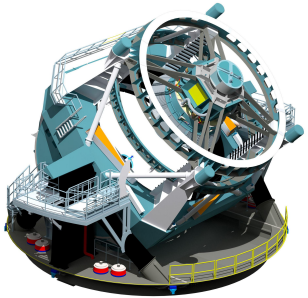
- Julien Peloton (Lead) - CNRS/IJCLab, IT department
- Chris Arnault (distributed computing) - CNRS/IJCLab, IT department
- Julius Hrivnac (database) - CNRS/IJCLab, IT department
- Marco Leoni (machine learning) - Université Paris-Saclay, IT department
- Sacha Pateyron (cloud management, deployment, monitoring) - CNRS/IJCLab, IT department

Data path

We would like the interesting ones ...



Machine learning Association



10 million alerts per night…